

The Effect of Thinning and Superobservations in a Simple One-Dimensional Data Analysis with Mischaracterized Error

ROSS N. HOFFMAN

NOAA/Atlantic Oceanographic and Meteorological Laboratory, and Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida, and Atmospheric and Environmental Research, Lexington, Massachusetts

(Manuscript received 30 November 2017, in final form 28 January 2018)

ABSTRACT

A one-dimensional (1D) analysis problem is defined and analyzed to explore the interaction of observation thinning or superobservation with observation errors that are correlated or systematic. The general formulation might be applied to a 1D analysis of radiance or radio occultation observations in order to develop a strategy for the use of such data in a full data assimilation system, but is applied here to a simple analysis problem with parameterized error covariances. Findings for the simple problem include the following. For a variational analysis method that includes an estimate of the full observation error covariances, the analysis is more sensitive to variations in the estimated background and observation error standard deviations than to variations in the corresponding correlation length scales. Furthermore, if everything else is fixed, the analysis error increases with decreasing true background error correlation length scale and with increasing true observation error correlation length scale. For a weighted least squares analysis method that assumes the observation errors are uncorrelated, best results are obtained for some degree of thinning and/or tuning of the weights. Without tuning, the best strategy is superobservation with a spacing approximately equal to the observation error correlation length scale.

1. Introduction

Weather forecasts depend on the accuracy of initial state estimates. A data assimilation (DA) system estimates the initial state or analysis by combining all recent observations with a prior estimate called the background (Kalnay 2002). The background is the analysis from the previous DA cycle propagated to the current time by a numerical model. Therefore, analyses of the initial state depend on the accuracy of the numerical model and of the characterization and specification of background and observation errors. In reality, background and observation errors are complex. Further, within the DA context, estimated observation errors include contributions from instrument error due to noise, systematic errors, and biases; representativeness error due to scales not included in the DA system; and simulation error due to inaccuracies in interpolation and in the calculation of the sensor response. Much progress has been made in using ensembles to estimate background errors, however, in practical DA systems,

observation error characterizations are often simplifications. Current DA systems often treat the observation errors as uncorrelated and employ some combination of thinning the observations (i.e., using a subset of available observations; e.g., Dando et al. 2007), creating superobservations (i.e., averaging subsets of available observations; e.g., Benjamin 1989), and tuning the estimated observation error standard deviations or observation weights (e.g., Li et al. 2009). These methods are suboptimal, but the impacts on analysis errors can be mitigated with careful tuning. However, the “optimal” tuning for one data sample may not be optimal for another since the true error statistics vary in space and time.

A number of authors have described various approaches to thinning, including Purser et al. (2000), Ochotta et al. (2005), Ramachandran et al. (2005), and Lazarus et al. (2010), and its impact on DA, including Dando et al. (2007), Li et al. (2010), and Miyoshi and Kunii (2012). Liu and Rabier (2002) reported on idealized data analyses in a one-dimensional (1D) periodic domain and compared optimal and suboptimal schemes and demonstrated that thinning is justified when observation error correlations are neglected. Liu and Rabier (2003) extended this work to a much more realistic

Corresponding author: Ross N. Hoffman, ross.n.hoffman@noaa.gov

system. As a rule of thumb, they found that the optimal thinning reduced the maximum remaining error correlations to about 0.15–0.20. These studies and an earlier study by Bergman and Bonner (1976) showed there is little to gain from decreasing the spacing of observations to be much less than the correlation length scale, even if the correct observation covariances are used.

The use of estimates of vertical or interchannel observation correlations is now commonplace (e.g., Bauer et al. 2011; Campbell et al. 2017). However, estimates and implementations of horizontal observation error correlations have been lacking. This is now changing. For example, to allow more data to be used in the DA system, Waller et al. (2016) diagnose observation error correlations of Doppler radar radial winds following the approach of Desroziers et al. (2005). Other similar efforts include those of Nakabayashi and Ueno (2017) and Hotta et al. (2017). Since including a nondiagonal observation covariance matrix \mathbf{R} entails a higher computation cost, especially when assimilating high-density satellite observations, Abdelnur Ruggiero et al. (2016) developed a method that requires only a block diagonal error correlation matrix by augmenting the observation vector with spatial derivatives of the observations.

In addition to higher computational costs, a full matrix \mathbf{R} also introduces the complication that typical observation error covariances often result in ill-conditioned system of equations. Therefore, \mathbf{R} must be accurately estimated. However, there are fundamental limits when estimating error statistics related to sample size and the fact that the differences available to be used in the estimation process combine errors from all sources (Todling 2015a). As a result, 1) errors can only be estimated in some (quasi) homogeneous sense and/or for situations that are well observed and 2) estimating one component of error requires assuming knowledge of the other components (Todling 2015b). Consequently, in practice, an excellent global estimate of an observation error covariance matrix may perform poorly for specific locations and times where the true error statistics differ from the overall statistics. For similar reasons, the optimal degree of thinning or superobservation and the optimal values for the observation weights actually vary in space and time. Therefore, several authors note that thinning should be adaptive. For example, Dando et al. (2007) state that “localized regions of the atmosphere containing large gradients such as frontal regions may benefit from [smaller] thinning distances ... and therefore the global optimal separation distance is not necessarily applicable in these circumstances.” Examples of such “intelligent” data selection include a tool for dynamically optimizing the thinning of satellite data (Zhu and Boukabara 2015)

and an efficient thinning method based on support vector regression (Richman et al. 2015).

All studies of the impacts of the various accommodations that make DA systems practical have limitations and caveats. In different realistic cases, different researchers have reported that different thinning or superobservations strategies are optimal for a particular problem. Comparisons across studies are hindered when the impacts due to one factor (e.g., thinning) are examined in one case and the impacts due to a second factor (e.g., observation error inflation) are examined in another case. The present study also has limitations. However, unlike the previously cited studies, the expected analysis error is directly calculated instead of estimated from a finite sample. It is common to calculate the expected analysis error under the assumption that the statistics used are correct, but it is likewise possible, as will be shown, to calculate the expected analysis error when the statistics used are incorrect. This approach enables a more comprehensive exploration of the effects of thinning and superobservation as the true and estimated error statistics vary. In this study, different thinning and superobservation methods are compared for varying degrees of thinning or superobservation for cases in which the estimated error characteristics are incorrect. Both correlated and systematic errors are considered. Systematic errors may occur when some relevant geophysical phenomena are not included in the observation simulation or when there are slowly varying observing system errors. Systematic errors are often termed biases (Dee 2005), but the systematic errors considered here have an expectation of zero and will not be referred to as biases. Two methods are considered: a variational analysis method that includes a full observation error covariance matrix \mathbf{R} and a weighted least squares analysis method in which it is assumed that \mathbf{R} is diagonal. The effect of variations in both \mathbf{R} and the diagonal weight matrix are examined. The approach is general and might be applied to arbitrary geometries and observation operators. However, examples presented are 1D and the observation operator is the simplest possible since the observation vector is a subset of the analysis vector.

The plan of this paper is the following. Section 2 presents the mathematical methodology employed. In section 2a, the expected errors of a variational analysis are determined under general conditions, assuming only that the background and observation errors are independent. Indeed, it is assumed that the background and observation error covariances are *not* correctly specified. In section 2b, the error covariance matrices are parameterized in terms of a small number of parameters, including the magnitude of random, correlated, and constant error components. All correlations

are based on a Gaussian hill shape, specified in terms of an e -folding length scale. Section 2c extends the results of sections 2a and 2b to the superobservation case. Section 2d presents the diagonal version of \mathbf{R} in the weighted least squares analysis method where observation errors are assumed to be uncorrelated, but the observation weights are tuned. The methodology is then applied to a very simple 1D univariate analysis problem (section 3). In the geometry of this problem the analysis locations are identical to the original (i.e., before thinning or superobservation) observation locations (section 3a). Then, the various observing networks (ONs; section 3b) are realized by very simple observation operators (section 3c). Section 4 presents some sample results, first for the variational analysis method (section 4a) and then for the weighted least squares analysis method (section 4b). Finally, a discussion of the results and concluding remarks are given in section 5.

2. Methodology

a. Expected errors of variational analysis

The variational analysis \mathbf{x}^a is the \mathbf{x} that minimizes the functional

$$J(\mathbf{x}) = \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \delta \mathbf{y}^T \mathbf{R}^{-1} \delta \mathbf{y}. \quad (1)$$

Here, \mathbf{B} is the estimated background error covariance, \mathbf{R} is the estimated observation error covariance, and

$$\begin{aligned} \delta \mathbf{x} &= \mathbf{x} - \mathbf{x}^b = \mathbf{e}^a - \mathbf{e}^b, \\ \delta \mathbf{y} &= \mathbf{y}^o - [\mathcal{H}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b)] = \mathbf{e}^o - \mathbf{H} \mathbf{e}^a, \end{aligned} \quad (2)$$

where \mathbf{x} is the state, \mathbf{x}^b is the background, \mathbf{y}^o is the observation vector, \mathbf{e} is the error, \mathcal{H} is the observation operator, and \mathbf{H} is its linearization. Here, the observation error \mathbf{e}^o includes both instrument and representativeness errors. (See the appendix for details.) There are N analysis locations in \mathbf{x} and n observations in \mathbf{y} . The i th observation is denoted y_i .¹ An assumption is often made in deriving Eq. (1) that the background and observation errors are Gaussian, or have been transformed to be Gaussian (e.g., Beal et al. 2010).

As shown in the appendix, the expected analysis error covariance matrix is given by

$$\hat{\mathbf{A}} = \mathbf{G} \hat{\mathbf{B}} \mathbf{G}^T + \mathbf{K} \hat{\mathbf{R}} \mathbf{K}^T = \hat{\mathbf{A}}^b + \hat{\mathbf{A}}^o, \quad (3)$$

where

$$\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T)^{-1} \quad (4)$$

and

$$\mathbf{G} = \mathbf{I} - \mathbf{K} \mathbf{H}. \quad (5)$$

The matrices \mathbf{K} and \mathbf{G} are the Kalman filter gain and covariance update matrices. Equation (3) defines the background and observation components of $\hat{\mathbf{A}}$ (denoted by superscripts b and o) due to the true background and observation covariances, $\hat{\mathbf{B}}$ and $\hat{\mathbf{R}}$, respectively.² The diagonal of $\hat{\mathbf{A}}$, denoted Δ_A , is the expected analysis error variance at the analysis locations. The average trace of $\hat{\mathbf{A}}$ and its components are the total, background, and observation MSEs, defined by

$$\varepsilon_A = \text{tr}(\hat{\mathbf{A}})/N = \text{tr}(\hat{\mathbf{A}}^b)/N + \text{tr}(\hat{\mathbf{A}}^o)/N = \varepsilon_A^b + \varepsilon_A^o. \quad (6)$$

A numerical recipe to determine $\hat{\mathbf{A}}$ given \mathbf{R} , \mathbf{B} , \mathbf{H} , $\hat{\mathbf{R}}$, and $\hat{\mathbf{B}}$ that requires only a single matrix inverse, first calculates \mathbf{K} from Eq. (4) and \mathbf{G} from Eq. (5), and then obtains $\hat{\mathbf{A}}$ from Eq. (3). For a fixed ON and estimated statistics, \mathbf{K} and \mathbf{H} are fixed. When this holds, if the true background (observation) covariance is changed, then only the background (observation) components of $\hat{\mathbf{A}}$, Δ_A , and ε_A will change.

b. Parameterized error covariances

To make use of the formulation of section 2a requires the specification of \mathbf{B} , \mathbf{R} , \mathbf{H} , $\hat{\mathbf{B}}$, and $\hat{\mathbf{R}}$. Specification of \mathbf{H} for thinning and superobservation is provided in section 3c. This section defines a generic parameterization of a covariance matrix \mathbf{P} that can be used to create all four of the needed covariance matrices.

To specify a particular model for the errors that allows the calculation of \mathbf{P} from a handful of parameters, it is assumed that \mathbf{e} is composed of three additive independent error sources according to

$$\mathbf{e} = \mathbf{q} + \mathbf{r} + e_c \mathbf{j}. \quad (7)$$

The uncorrelated error component \mathbf{q} is a multivariate random variable with zero mean and covariance $\sigma_q^2 \mathbf{I}$. The correlated error component \mathbf{r} is a multivariate random variable with zero mean, constant variance σ_r^2 , and correlation

$$\rho_{ij} = e^{-D_{ij}^2/\lambda^2}, \quad (8)$$

¹This typographical convention is used for elements of other vectors and matrices.

²The term covariance matrix is used throughout this paper. However, no assumption was made that the errors are unbiased. Therefore, it would be more correct to use the term expected cross-product matrix for $\hat{\mathbf{B}}$ or $\hat{\mathbf{R}}$.

where D_{ij} is the distance from location i to location j and λ is the correlation length scale. The systematic error component $e_c \mathbf{j}$ is a “constant” that is added to all observations for each specific realization. Here, e_c is a scalar random variable with zero mean and variance σ_c^2 , and \mathbf{j} is a vector of ones. Since the three components are independent,

$$\begin{aligned} \mathbf{P} &= \langle \mathbf{e} \mathbf{e}^T \rangle \\ &= \langle \mathbf{q} \mathbf{q}^T \rangle + \langle \mathbf{r} \mathbf{r}^T \rangle + \langle e_c^2 \rangle \mathbf{j} \mathbf{j}^T \\ &= \sigma_q^2 \mathbf{I} + \sigma_r^2 \boldsymbol{\rho} + \sigma_c^2 \mathbf{J}. \end{aligned} \tag{9}$$

Here, \mathbf{J} is a matrix of ones. Thus, \mathbf{P} is defined once the D_{ij} are calculated and the parameters σ_q , σ_r , σ_c , and λ are specified. In what follows each of these parameters may have a superscript b or o to indicate background or observation and a hat accent to indicate true as opposed to estimated. Note that the uncorrelated and systematic error components may be considered additional (but independent) correlated error components with zero and infinite correlation length scales, respectively.

c. Superobservations

Superobservations (aka superobs) are simple averages of the observations. The location of a superobservation is the mean of the locations of the original observations that were averaged. To specify a method to create superobservations, for each of the n superobservations, y_i , give a list K_i of the indices k of the original observations y'_k in the i th superobservation. Let n' be the number of original observations and let m_i be the number of indices in K_i . (The thinning case may be considered a specialization of the superobservation case where $m_i = 1$.) With K_i given, the superobservation average can be written as

$$y_i = \frac{1}{m_i} \sum_{k \in K_i} y'_k, \tag{10}$$

where the notation indicates that the sum is over the k in the list K_i . The superobservation average applies to the location, observed value, and observation error.³ Equation (10) can be written in matrix–vector notation as

$$\mathbf{y} = \mathbf{S} \mathbf{y}'. \tag{11}$$

Here, \mathbf{S} is the selection or superobservation matrix. For the thinning case \mathbf{S} is just an indicator matrix that is all zeros, except that in the i th row the k th entry is 1, where y'_k is taken to be the k th original observation. For the

superobservation case, \mathbf{S} is a normalized indicator matrix that is all zeros, except that in the i th row the m_i entries given by K_i are equal to $1/m_i$. In other words the superobservation operator for y'_i is used to select the original observations for k in K_i and average them.

All of the developments of section 2a apply equally to observations or superobservations. Given a set of original observations \mathbf{y}' and K_i , the above description allows the calculation of m_i . Then, the observations \mathbf{y}^o and their locations are determined from Eq. (10) or (11). In some usages, the observation operator for superobservations is the same as the observation operator for a single original observation. However, small scales present in reality are filtered by the superobservation averaging, and an alternative is to include this averaging in the observation operator so that small scales in the background are similarly filtered. This alternative approach is used in the examples presented below. All that remains is to determine the error covariance matrices for the superobservations from the error covariance matrix of the original observations. Let \mathbf{P} denote either \mathbf{R} or $\hat{\mathbf{R}}$ and let \mathbf{P}' denote the corresponding original covariance matrix. Then, elements of \mathbf{P} are simply averages of all elements of \mathbf{P}' that correspond to the original observations averaged over because the superobservation errors are formed from the original observation errors according to Eq. (10); that is,

$$\begin{aligned} P_{ij} &= \langle e_i e_j \rangle \\ &= \frac{1}{m_i} \frac{1}{m_j} \sum_{k \in K_i} \sum_{l \in K_j} \langle e'_k e'_l \rangle. \\ &= \frac{1}{m_i} \frac{1}{m_j} \sum_{k \in K_i} \sum_{l \in K_j} P'_{kl} \end{aligned} \tag{12}$$

Equivalently,

$$\mathbf{P} = \langle \mathbf{e} \mathbf{e}^T \rangle = \mathbf{S} \langle \mathbf{e}' \mathbf{e}'^T \rangle \mathbf{S}^T = \mathbf{S} \mathbf{P}' \mathbf{S}^T. \tag{13}$$

d. Weighted least squares

In practice, and in the examples that follow, it is often assumed that the observation errors are uncorrelated and unbiased. In that case, \mathbf{R} is diagonal and it is possible to write Eq. (1) as

$$J(\mathbf{x}) = \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \sum_i^n w_i \frac{[y_i^o - \mathcal{H}_i(\mathbf{x})]^2}{\sigma_o^2}, \tag{14}$$

where σ_o^2 is the estimated variance of the observation errors and the w_i are ad hoc or tuning weights. If the observation errors are in fact uncorrelated and unbiased

³ This assumes that the observation errors are additive.

with standard deviation $\hat{\sigma}_q^o = \sigma_o$, then the optimal choice of weights is $w_i = 1$. But if the observation error characteristics are different, then other values of the w_i will be optimal. Also, in cases where the observations are superobservations of varying numbers (m_i) of original observations with observation errors that are uncorrelated and unbiased with standard deviation $\hat{\sigma}_q^o = \sigma_o$, then $w_i = m_i$ will be optimal weights. Equation (14) is equivalent to Eq. (1) with

$$\mathbf{R} = \sigma_o^2 \mathbf{W}^{-1}, \quad (15)$$

where \mathbf{W} is a diagonal matrix with the w_i along the diagonal.

3. A simple 1D data analysis problem

While section 2 is fairly general, a simple 1D data analysis problem is now defined that is appropriate for investigating the impact of thinning and superobservation when the error statistics are misspecified. In this setup, the analysis variables and the observations before thinning or superobservation are identical. That is, the locations are the same and the physical quantity analyzed and observed is the same (e.g., temperature). As a consequence, $n' = N$ and the observation operator for the original observations is the identity matrix. The analysis locations are denoted h_i , which could be the height above the ground, or the log of pressure, or the distance along a horizontal line. In the discussion that follows h is taken to be the horizontal distance across a swath of satellite observations relative to the center of the swath, but the results may be applied to other domains with the caveat that the correlations depend only on distance and not location within the domain according to Eq. (8). The background and observation, estimated, and true errors are all potentially of the form of Eq. (7): the sum of random uncorrelated, random correlated, and systematic error components, all independent of each other and the background. However, the solution of the analysis problem may assume different values for the parameters defining the error statistics. The examples in section 4b use the weighted least squares approach [\mathbf{R} is given by Eq. (15)].

The simple 1D analysis problem may be usefully extended with relatively small changes. First, if h is a vertical dimension such as the logarithm of pressure, this 1D problem is of interest for testing thinning/superobservation strategies for assimilating temperature profiles retrieved from radiance or radio occultation measurements. Second, the analysis could be extended to radiances or bending angles by using an appropriate observation function. Third, to extend the 1D analysis problem to higher dimensions,

TABLE 1. Nominal parameter values used in Eqs. (8) and (9). See the text for the definition of the symbols.

\mathbf{P}	σ_q	σ_r	σ_c	λ
\mathbf{B}	0	1	0	16
\mathbf{R}	0.1	1	0	8
$\hat{\mathbf{B}}$	0	1	0	16
$\hat{\mathbf{R}}$	0.1	1	0	8

it is only necessary to formulate the thinning and superobservation strategies in terms of distance, appropriately defined.

a. Setup

To be concrete, consider a cross section through a swath of satellite data, with y'_k the original observations before thinning or superobservation at the k th location. In the examples in section 4, k varies from 1 at the left edge of the satellite track to $N = n' = 65$ at the right edge with $k = 33$ at nadir, and h varies from -32 at the left edge to $+32$ at the right edge with $h = 0$ at nadir. In physical units, if the distance between adjacent observations is $\delta h = 25$ km, then the center-to-center swath width is 1600 km. Table 1 lists the parameters for the nominal case, which corresponds to true and estimated background and observation errors that are correlated, with the background correlation length scale equal to twice the observation correlation length scale, and with no systematic errors.

b. Observing networks

Thinning of the observations keeps every m th location. Patterns of superobservations average every m observations. The parameter m is the *degree* of thinning or superobservation. The thinning and superobservation ONs that will be used in the examples are plotted in Fig. 1.

- The “odd” thinned ONs include the central location and then every m location in both directions, for m odd. The locations for the odd thinned and superobservation ONs of the same degree are identical. (Compare the ochre⁴ and black locations in Fig. 1.) In the odd superobservation ONs the superobservation averages are over the m original observations closest to each superobservation location. In what follows, only values of m that are factors of 63 are used, the edge locations (mint locations in Fig. 1) are not used, and each odd superobservations ON is based on the remaining 63 original observations

⁴ Colors are from the Dark2 color-blind safe palette of ColorBrewer: mint, ochre, lavender, magenta, lime, gold, and brown.

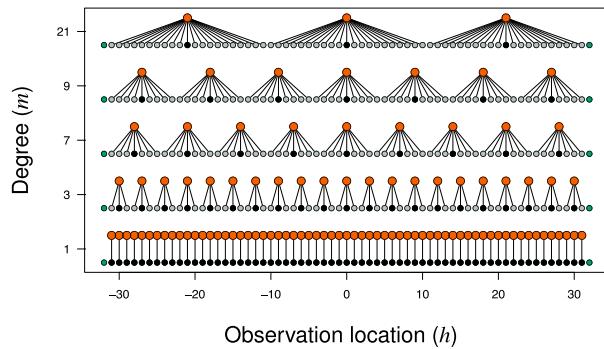


FIG. 1. Thinning and superobservation ONs used in the examples. The superobservation locations are shown by large ochre circles and the thinned locations by black circles. Original locations (gray or black circles) contributing to a superobservation location are connected by lines to that superobservation. The original edge locations added to the odd ONs to create the hybrid ONs are shown by mint circles. All ONs used are symmetric and all include the central location.

- The “hybrid” thinned and superobservation ONs are identical to the corresponding odd ONs but add the edge locations (mint locations in Fig. 1). These ad hoc hybrid ONs are motivated by a desire to retain information for the full range of h . However, the hybrid superobservation ONs have the additional complexity that these are mixtures of superobservations and original observations and that the original observation locations skipped between the edge location and next location are only half of the usual number $[(m-1)/2]$ instead of $(m-1)$. In the nominal case the weight given to edge locations is 1 and the weight given to the superobservation locations is m .

Since analyses depend strongly on the locations of the observation, these ONs are designed so that some characteristics of the ONs are fixed in certain comparisons. Examples of this include the following. Observation locations are the same for thinned and superobservation ONs of the same degree. A pair of similar odd and hybrid ONs differ only in that the edge locations are included in the hybrid ON. Superobservation ONs of different degrees include the same set of original observations.

c. Observation operators

The observation operators for the setup and ONs considered here are extremely simple. Since the observed and analyzed quantity are identical and since the original observation locations are the same as the analysis locations, the observation operator is equal to the selection/superobservation operator (i.e., $\mathbf{H} = \mathbf{S}$). In other words the observation operator for y_i^o is to select the analysis values x_k^a for k in K_i and average them. (See section 2c.)

d. Numerical solution

Determining the MSE requires inverting the matrix $\mathbf{C} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$ in Eq. (4). This matrix can be poorly conditioned or even numerically singular. In the results presented in section 4, $\sigma_q^o = 0.1$. This adds a ridge to \mathbf{R} of 0.01 and tends to stabilize the solution. However, in some cases, especially for $m = 1$ and 3, this is not sufficient and a generalized inverse is used in all calculations. To avoid anomalous behavior, a maximum condition number of 100 is enforced. Typically, the leading 10 or so singular vectors are retained, which account for more than 99% of the original variance. For example, for the nominal case and the $m = 1$ odd ON, 10 singular vectors (out of 63) are kept, but only 0.9% of the original variance is removed.

4. Results

a. Sensitivity of analysis error to misspecification of uncertainty

When the specified background and observation covariances are correct, \mathbf{x}^a is the optimal analysis (OA) and the expected MSE calculated from Eq. (6) will be minimized. Since it is impossible to know the truth, it is likewise impossible to know the errors exactly. As a result, the variational analysis specification of the covariances will never be perfect. To examine the impact on the analysis error when this happens, Fig. 2 plots the expected analysis error as the estimated statistic parameters defining \mathbf{B} and \mathbf{R} are varied. Note that the y axes in Fig. 2 are identical and that the x axes are identical for each parameter normalized by its true value. The left panels in Fig. 2 examine what happens if the estimated standard deviations are misspecified by varying σ_r^b and σ_r^o about their nominal value (of 1) for the odd ONs. Increasing the degree m results in increasing MSE. Superobservation is an improvement compared to thinning, but this improvement is small for all cases except for $m = 21$ since the superobservation average is over distances approximately equal to or smaller than the observation error correlation length scale. For $m = 1$, the analysis is identical for thinning and superobservation, but MSE increases faster for thinning than superobservation as m increases. The extreme of $m = 21$ is substantially worse than other choices. Here and in what follows, the impact of m on superobservation results is muted because the superobservation process retains all the information in the original observation and filters the errors. These findings are consistent with the theoretical conclusions of Xu (2011) for uniformly distributed observations and with the numerical results of Xu and Wei (2011) for radar observations of radial velocity. Since Eq. (10) applies to errors, this filtering of errors works best for uncorrelated

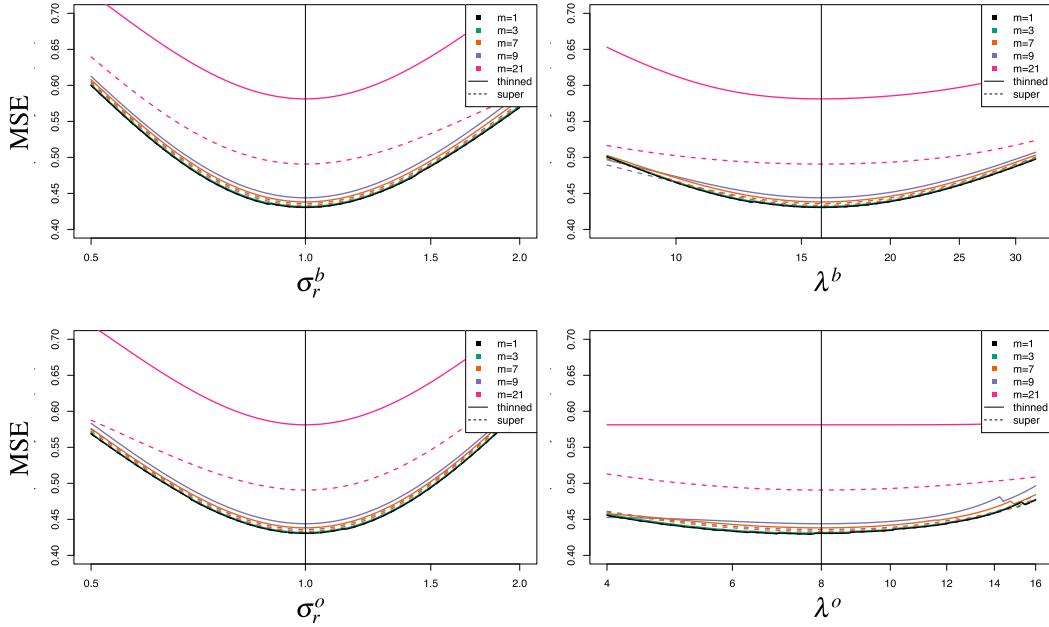


FIG. 2. The variational analysis MSE in the thinning case (solid lines) and in the superobservation case (dashed lines) for the odd ONs as a function of perturbing the estimated statistic parameters—(left) σ_r , and (right) λ for both (top) **B** and (bottom) **R**—from 0.5 to 2.0 times the true values, for different values of the thinning parameter (m , colors).

error and has no impact on systematic errors or bias. In the superobservation case, while the same set of original observations contribute for each m , as m increases the ON locations collapse toward the center of the domain, which has a negative impact on the analysis error.

Increases in the analysis error due to misspecification of the estimated standard deviations are large compared to the variations in analysis error due to the degree of thinning or superobservation (except for the case $m = 21$) and due to the choice of thinning or superobservation, as seen by the close packing of the curves in Fig. 2. Results for varying σ_r^b are qualitatively similar to results for varying σ_r^o . The right panels in Fig. 2 are analogous to the left panels but for varying λ^b and λ^o . The sensitivities to percentage errors in the error correlation length scales are half or less than those in the error standard deviations.

Another approach to examining the sensitivity of analysis error to deficiencies in the estimated error model is to hold the estimated error model fixed and vary the true errors. This is of practical interest in cases where a fixed global error model is used, but the true errors vary in space and time. Because of the form of Eqs. (3), (6), and (9), it is possible to consider the impact of the different parameters separately. Substituting Eq. (9) for the true covariances in Eq. (3), it is clear that the variance parameters such as $\hat{\sigma}_r^{o2}$ appear only as a multiplier of a single term. That term is therefore the sensitivity of the ε_A with respect to that variance parameter. For example,

$$\frac{\partial \varepsilon_A}{\partial \hat{\sigma}_r^{o2}} = \text{tr}[\mathbf{K}\boldsymbol{\rho}(\hat{\lambda}^o)\mathbf{K}^T]/N. \tag{16}$$

These sensitivities are plotted in Fig. 3. Note that the terms in Eq. (9) involving σ_q^2 and σ_c^2 are equal to the term involving σ_r^2 for λ equal to zero and ∞ , respectively. The curves in Fig. 3 are at the same time a plot of the variation of ε_A^b and ε_A^o with respect to $\hat{\lambda}^b$ and $\hat{\lambda}^o$, respectively, holding the other parameters constant. Note that for the OA, ε_A^b and ε_A^o are approximately equal to 0.2. Considering the top panel, as $\hat{\lambda}^b$ decreases, ε_A^b increase to $O(1)$. On the other hand, as $\hat{\lambda}^b$ increases from its nominal value of 16, ε_A^b decreases slightly. In contrast, in the bottom panel in Fig. 3, as $\hat{\lambda}^o$ decreases, ε_A^o decreases slightly for the larger m for the thinned ONs and decreases very substantially for $m = 3$ and to near zero for $m = 1$ and all the superobservation ONs. Finally, as $\hat{\lambda}^o$ increases, ε_A^o approximately doubles. The conclusion is that if everything else is fixed, the analysis error decreases with increasing true background error correlation length scale and with decreasing true observation error correlation length scale. This matches a priori expectations based on an information content point of view. Since the degrees of freedom decrease with increasing correlation length scales, there is less information needed to correct a background with a long correlation length scale and more information to do so in observations with a short correlation length scale.

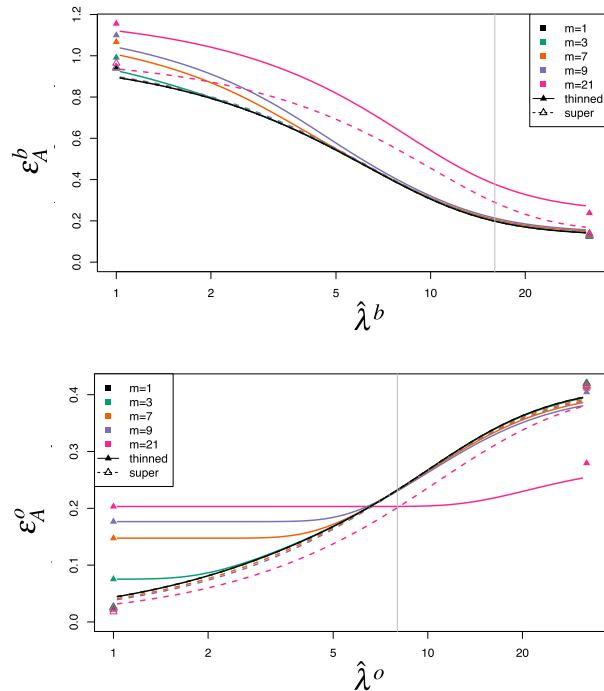


FIG. 3. The variational analysis mean square (top) background (ϵ_A^b) and (bottom) observation (ϵ_A^o) analysis error for the nominal parameters in the thinning case (solid lines) and in the superobservation case (dashed lines) for the odd ONs as a function of perturbing the true correlation length scales—(top) $\hat{\lambda}^b$ and (bottom) $\hat{\lambda}^o$ —for different values of the thinning parameter (m , colors). Symbols plotted at the left and right are for $\lambda = 0$ and ∞ , respectively.

b. Tuning the weighted least squares analysis method

With a given analysis system that may be imperfect because of mischaracterization of the observation errors, it is common to attempt to optimize performance by tuning the observation weights or estimated observation error standard deviation, by thinning the data, and/or by generating superobservations. In this section, for some examples using the least squares approach, the impact on the estimated MSE of the analysis is calculated for different degrees m of both thinning and superobservation, and optimal values of the weights w_i are determined. In this section MSE is calculated using Eq. (6) specialized with Eq. (15), and MSE is minimized with respect to w_i using the standard R nonlinear minimization software nlm.

Figure 4 demonstrates the need for tuning the data thinning, superobservation, and/or weighting for the success of the weighted least squares analysis method. Here, $\sigma_o = 1$ and $w_i = m_i$. For reference, the OA MSE is 0.4270 for the hybrid $m = 1$ ON (i.e., for the best use of all the observations) and 0.4305 for the odd $m = 1$ ON. Since the optimal values of σ_o are approximately 1.1 and 1.3 for $m = 9$ and $m = 7$, if σ_o is not tuned ($\sigma_o = 1$), then

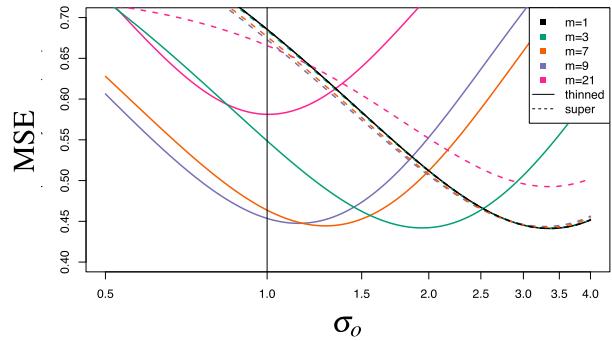


FIG. 4. The weighted least squares MSE in the thinning case (solid lines) and in the superobservation case (dashed lines) for the odd ONs as a function of perturbing the estimated observation error— σ_o —from 0.5 to 4.0 times the nominal value, for different values of the thinning parameter (m , colors).

simple thinning with a choice of $m = 9$ is best, and a choice of $m = 7$ is only a bit worse. For these choices the spacing between thinned observations is close to the true observation correlation length $\hat{\lambda}^o = 8$, and the closest-neighbor correlations are 0.28 and 0.47, respectively. Using all the observations either directly or in superobservations without tuning of the weights yields MSE values in excess of 0.65. This poor performance is the result of oversampling what are essentially redundant observations. Either thinning or inflating the estimate of σ_o ameliorates this poor performance. Thus, in Fig. 4 for each ON there exists an optimal choice of σ_o (equivalent to an optimal choice of w) that gives an MSE of approximately 0.45. For example, all the superobservation ONs, except for $m = 21$, are skillful for σ_o in the range of 3–4.

Three ways of tuning the weights in the weighted least squares analysis method were tested:

- w1—The w_i are specified by a single weight; that is, the weights are the same, except that in the hybrid superobservation case the edge weights (i.e., the weights for the observations at the edge locations) are reduced by a factor of $1/m$.
- w2—The w_i are specified by two weights; that is, the edge weights vary independently of the other locations.
- wn—The w_i vary independently.

Since the ONs are symmetric about $h = 0$, the weights should be as well. If unconstrained, the wn minimization finds weights that are very close to symmetric, but some weights are found to be negative. In what follows, the weights are constrained to be positive and symmetric by choosing the control vector for the minimization as the square root of the weights for $h \geq 0$. The three ways of tuning the weights were applied to the four ONs—odd

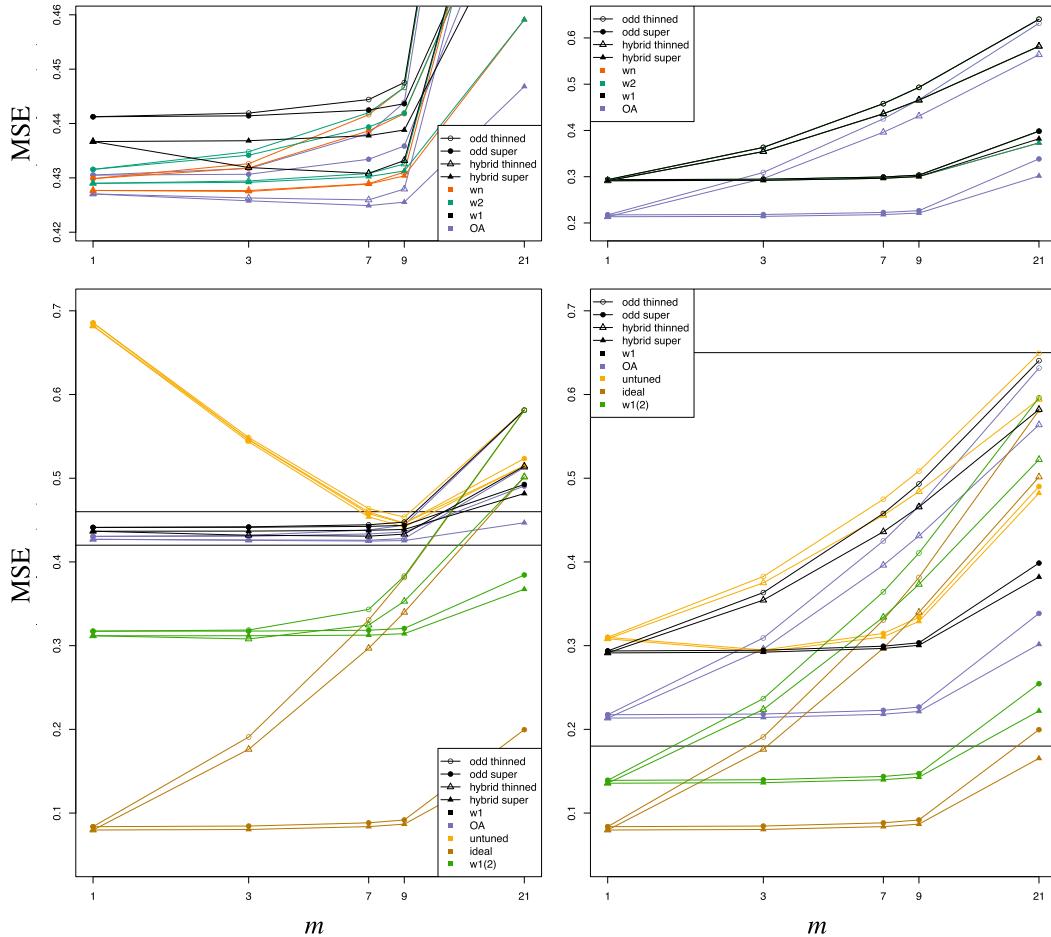


FIG. 5. The optimal MSE as a function of the degree of thinning (m) for (top) the three methods of tuning (colors) for the four ONs (thinned and superobservation, odd and hybrid, symbols) for (left) the nominal case ($\hat{\lambda}^o = 8$) and (right) the case with systematic errors ($\hat{\sigma}_c^o = 0.50$). For reference, lavender lines show the OA MSE for the four ONs. Note the difference in the y axis in the top panels. (bottom) For context, the w1 and OA results from the top panels are repeated in the inset defined by the black horizontal lines using the same y axis and the untuned ($w = 1$) MSE (gold) and the MSE are added for the case of ideal (uncorrelated and unbiased) errors (brown) for the four ONs. In addition, a second tuning case is added in each bottom panel as a second and lower set of w1 lines (lime)—for (left) correlated errors ($\hat{\lambda}^o = 4$) and for (right) systematic errors ($\hat{\sigma}_c^o = 0.25$).

and hybrid, thinned, and superobservation. It is easy to justify w2 tuning for the hybrid super-ONs. However, since it was found in many cases, in contrast to a priori expectations, that the optimal edge weight is larger than for the superobservation locations, w2 tuning was applied to all ONs. Further, since in an optimal analysis system the effective weights applied to each observation are different, the wn tuning was also tested.

A summary of results for the optimal analysis errors is displayed in Fig. 5 for two test cases. In the case of correlated errors (nominal parameter values, left panels), adding degrees of freedom in the weight tuning (going from w1 to w2 to wn) improves the analysis accuracy in general, but this improvement decreases as m

increases, except for the hybrid superobservation ONs. Increasing m increases MSE, except for the hybrid w1 thinned ONs. In the case of systematic errors added to uncorrelated errors (nominal parameter values except that $\sigma_q = 1$, $\sigma_r = 0$, and $\sigma_c = 0.5$ for \mathbf{R} , right panels), there is essentially no improvement by adding degrees of freedom in the weight tuning, and almost all of the w2 and wn points are overplotted by the w1 points. Additional general conclusions in both cases are that the observations at the edge locations add value, and that superobservation generally is better than thinning. For comparison the top panels in Fig. 5 include in lavender the OA MSE values. The bottom panels repeat the top panels, but without the w2 and wn results and expand

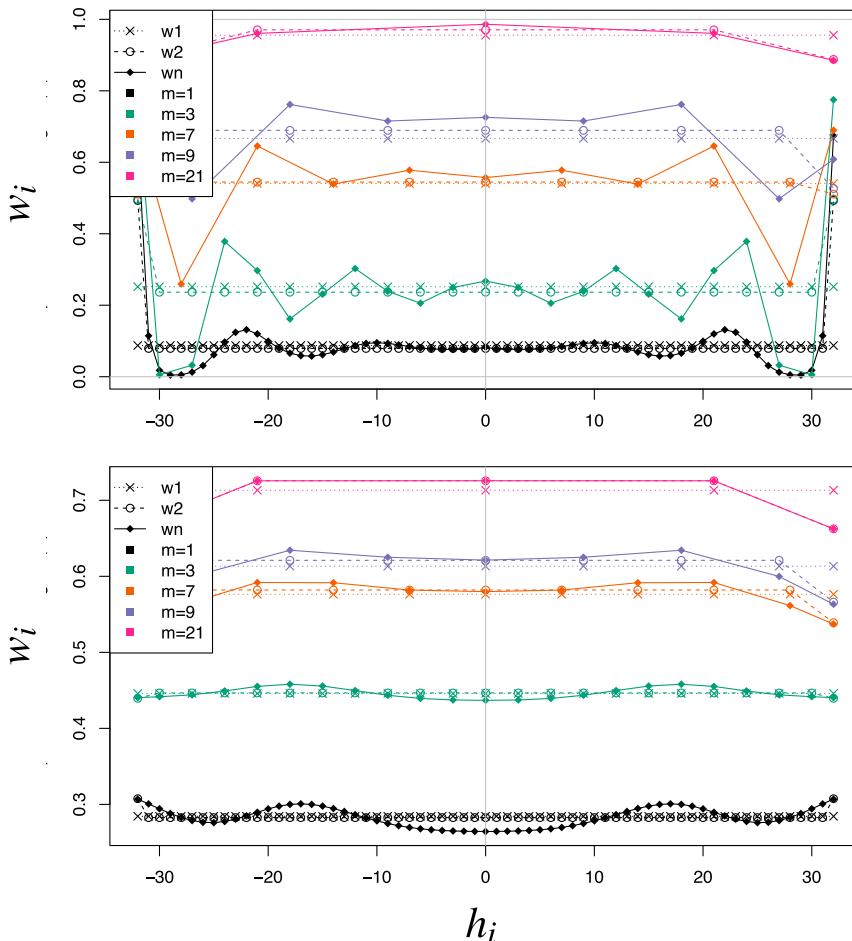


FIG. 6. The optimal weights w_i as a function of observation location h_i for representative cases for different degrees (m , colors, as in earlier figures) and for different weight optimizations (symbols and line types), for hybrid thinned ONs. The cases are for (top) correlated errors ($\lambda^o = 8$) and (bottom) observations with systematic errors ($\sigma_c^o = 0.50$). Note different ranges along the y axes.

the vertical axis to include the weighted least squares MSE values both without tuning of the weights (gold) and for the case of ideal uncorrelated unbiased errors (brown). Within this context, for correlated errors, any of the ONs is adequate as long as it is tuned in some way, and the tuned weighted least squares solutions are nearly as good as the OA solutions (using a full estimated covariance matrix \mathbf{R} equal to the true covariance matrix $\hat{\mathbf{R}}$). For systematic errors, superobservation is better than thinning; tuning the weights adds little, but the OA is an improvement. In this case, since most of the total error variance is due to the random uncorrelated errors ($\sigma_q = 1$ and $\sigma_c = 0.5$), superobservation is effective at filtering the random component of the error. To show the sensitivity of these results to the true statistics, each bottom panel includes an additional set of w1 results (lime) for $\lambda^o = 4$ (left) and for $\sigma_c^o = 0.25$ (right).

The pattern of optimal individual weights (Fig. 6) warrants a brief discussion. In many optimal weight patterns, when the errors are correlated, the outermost locations are given the largest weights (for both the w2 and wn schemes). The reason for the emphasis on observations further from the center of the ONs is that in the presence of correlations, the observations overall must be downweighted, but the observations at the domain edges are more valuable since at the edges there are fewer observations (i.e., there is a lower observation density). This is also the case for observations with systematic errors and odd ONs, but not for observations with systematic errors and hybrid ONs. The examples shown in Fig. 6, which are for hybrid thinned ONs, are otherwise representative of the types of behavior seen in other cases. In the top panel in Fig. 6 for the nominal case with correlated observation errors

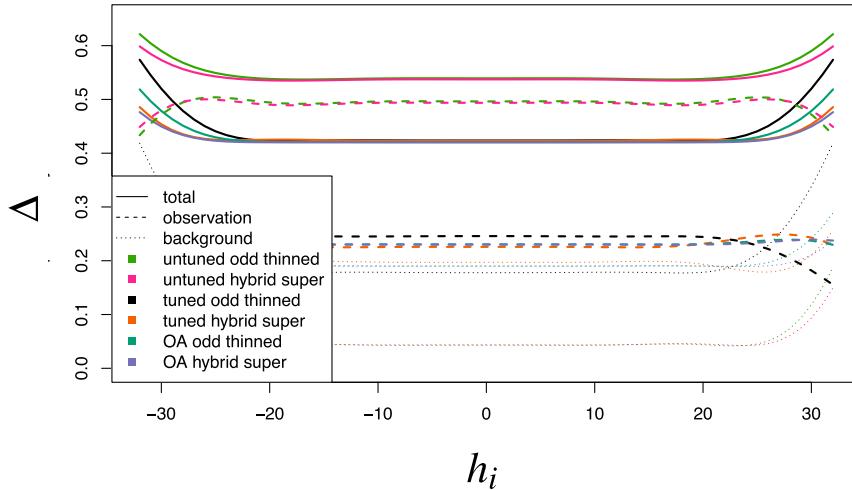


FIG. 7. The total and component squared analysis errors ($\Delta = \Delta_A, \Delta_A^b, \Delta_A^o$, line types) as a function of observation location (h_i) for degree $m = 3$ for the w1 thinned odd ON and for the w2 superobservation hybrid ON for the tuned and untuned weighted least squares solutions and for the OA solutions (colors; see legend) for the nominal case.

($\hat{\lambda}^o = 8$) the optimal observation weight patterns oscillate with increasing amplitude as $|h|$ increases, with large edge weights and in the case of $m = 1$ and 3 zero weights for some of the neighboring points. This provides an analog to thinning for $m = 1$ and 3 since the oscillation has a spatial wavelength equal to the correlation length scale. In the bottom panel in Fig. 6, for observations with systematic errors ($\hat{\sigma}_c^o = 0.50$) the optimal wn observation weight patterns are bent downward in the center and have a wavy pattern for increasing $|h|$. The edge weights are reduced in both the w2 and wn solutions, except for $m = 1$.

The pattern of squared analysis error (i.e., Δ_A , the diagonal of $\hat{\mathbf{A}}$) as a function of h is flat except within approximately $\hat{\lambda}^o = 8$ of the domain edge (Fig. 7). This is consistent with the fact that the background error is constant ($\hat{\sigma}_r^o = 1$) and the observation density is regular, except at the domain edge, where there is less information added by the observations. Except near the domain edges, the squared analysis error is virtually the same for all four of the tuned and OA cases. The background and observation components of squared analysis error— Δ_A^b and Δ_A^o —are also plotted in Fig. 7. At the domain edges, where the observation density is lower, the background component of squared analysis error (Δ_A^b) is higher because there are fewer observations to correct the background. Note that for some of the weighted least squares solutions, the observation component of squared analysis error (Δ_A^o) is lower at the domain edge when the edge location is not used or its weight is not tuned. However, in these cases, the impact of increasing Δ_A^b dominates.

5. Discussion and concluding remarks

A variational analysis problem is defined and analyzed to explore the interaction of observation thinning or superobservation with observation errors that are correlated or systematic. The expected errors of the variational analysis are determined under general conditions, assuming only that the background and observation errors are independent. The general formulation might be applied to arbitrary covariances, arbitrary observation operators, and arbitrary observation networks (ONs). For example, by comparing the impact of different estimated background error covariances, it would be possible to examine the impact of different implementations of covariance localization in ensemble data assimilation (DA). In this study, examples are given for a simplified analysis problem. In this problem, each covariance matrix is parameterized in terms of two parameters, the observation error standard deviation σ and the observation error correlation length scale λ . (This can include uncorrelated errors, in which case $\lambda = 0$, and systematic errors, in which case $\lambda = \infty$.) Also, only a regular 1D ON is considered in which the analysis variable and locations match the observation variable and locations. Then, the data selection operator \mathbf{S} and the observation operator \mathbf{H} are identical. In the discussion of the examples, it is assumed that the observation (and analysis) locations are separated by a fixed horizontal distance (25 km). However, the results are directly applicable to a vertical analysis problem with locations separated by a fixed log of pressure, say for thinning or superobservation of closely spaced temperatures retrieved from radiance or

radio occultation observations. Given an appropriate observation operator, a similar analysis could be made for thinning or superobservation of radiances or bending angle observations in order to develop a strategy for the use of such data in a full DA system.

In this study, different thinning and superobservation methods are compared for varying degrees of thinning or superobservation for cases in which the estimated error characteristics are incorrect. First, for the full variational analysis method, which includes an estimate of a full observation error covariance matrix \mathbf{R} , the analysis is more sensitive to variations in the estimated background and observation error standard deviations than in the corresponding correlation length scales. Except for extreme ONs, the impact of the ONs is small compared to the impact of the estimated statistical parameters. Since the superobservation ONs retain basically all the information in the original observations, as the degree of thinning or superobservation (m) increases, the mean square analysis error (MSE) increases more quickly for the thinned ONs than for the superobservation ONs.

Even if the estimated statistics are correct in a global sense, the true statistics will vary in space and time. In this situation, the expected analysis error may be locally larger or smaller than the global average, but it will always be suboptimal. If everything else is fixed, the analysis error increases with decreasing true background error correlation length scale and with increasing true observation error correlation length scale. This is expected since as the background degrees of freedom decrease with increasing background correlation length scales, there is less information needed to correct the background and as the observation degrees of freedom increase with decreasing observation correlation length scales, there is more information available to correct the background.

For a variety of reasons, \mathbf{R} is often assumed to be diagonal in practice. Then, in the case of correlated observation errors, thinning and/or superobservation and/or tuning of the observation errors improve analysis results. Therefore, in addition to results for the full variational analysis method, results are also presented for a weighted least squares analysis method in which \mathbf{R} is diagonal. In the weighted least squares analysis method, when the DA system ignores correlations and systematic errors, it is best to deweight the observations (i.e., tune the estimated observation standard deviation) or to create superobservations at the scale of the error correlation length scale. (With superobservations, the observation standard deviation of the original observations should be used without an $m^{-1/2}$ adjustment.) For errors with no systematic component, as long as the weights have been tuned, any of the ONs is adequate, and the tuned weighted least squares solutions are nearly as good as the

optimal analysis (OA) solutions. When systematic errors are added to uncorrelated errors, superobservation is better than thinning, but tuning the weights adds little. In this case, the OA solutions are a definite improvement, which suggests adding another parameter to the weighted least squares analysis method to estimate the systematic error in what [Dee \(2005\)](#) terms a “bias aware” approach. In the weighted least squares approach, the optimal configuration will also vary if the true error correlation length scales vary from sample to sample.

Two sets of ONs were examined that differ only in whether locations at the edge of the domain are included (hybrid ONs) or not (odd ONs). Differences in MSE are small between matching hybrid and odd ONs, with slightly smaller errors for the hybrid ONs since two additional observations are used and the range of observation locations is increased. Tuning of the observation weights illustrates the fact that within the weighted least squares context, when errors are correlated, the value of a particular observation depends on the location of the neighboring observations. Isolated observations or observations in regions of reduced data density have more impact on the analysis. The tuning gives increased weight to such observations.

Within the context of standard DA theory, it is often stressed how critical it is to properly characterize the observation errors, correct systematic errors, and estimate and use observation error correlations. As discussed in the introduction, this is difficult in practice because 1) limited samples of observations are available, 2) the correlations and systematic errors may vary with the particular batch of data, and 3) differences available for analysis combine multiple error sources. For example, for radiance observations, ignoring or mischaracterizing aerosols may induce systematic and/or correlated errors, and aerosols vary temporally and spatially on scales of interest.

The examples presented in this study demonstrate some of the complexity of optimizing a data analysis even for a very simple system. What can be done? In the near term, adaptive methods as described in the literature and cited in the introduction, hold some promise for enhancing analyses by sensibly retaining more observations as the feature length scales decrease. Various schemes to remove biases and systematic errors can be applied (e.g., [Dee and da Silva 1998](#)). And dense datasets can be preprocessed to more manageable sizes using smoothing splines or other interpolative techniques. Beyond standard DA practice, in the future artificial intelligence (AI) techniques based on deep learning (multilevel neural networks) might be trained on “big” data (initially simulated data) to produce analyses that are more nearly optimal for every case likely to occur in reality.

Acknowledgments. The author thanks his colleagues who encouraged and challenged him in this work. The reviewers provided valuable suggestions that improved this paper. This work was supported by NOAA (Grant NA15OAR4320064).

APPENDIX

Derivation and Discussion of Variational Analysis Expected Errors

The analysis \mathbf{x}^a occurs when $\partial J/\partial \mathbf{x}$ is zero; that is, when

$$\mathbf{B}^{-1} \delta \mathbf{x} = \mathbf{H}^T \mathbf{R}^{-1} \delta \mathbf{y}. \tag{A1}$$

In what follows, we solve for the analysis error, but note that solving Eq. (A1) for the analysis using Eq. (2) gives the Kalman filter update equation:

$$\delta \mathbf{x} = \mathbf{K}[\mathbf{y}^o - \mathcal{H}(\mathbf{x}^b)], \tag{A2}$$

where \mathbf{K} is defined below in Eq. (A8).

Now, the truth at the observation locations is given by

$$\mathbf{y}^t = \mathcal{H}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x}^t - \mathbf{x}^b) + \mathbf{e}^H, \tag{A3}$$

where \mathbf{e}^H is the combined errors of the observation operator (aka, forward problem) and its linearization. The truth in the DA context does not include so-called representativeness errors, which are the small scales that cannot be represented by the analysis. Using Eq. (A3),

$$\begin{aligned} \delta \mathbf{x} &= \mathbf{x}^a - \mathbf{x}^b = (\mathbf{x}^a - \mathbf{x}^t) - (\mathbf{x}^b - \mathbf{x}^t) = \mathbf{e}^a - \mathbf{e}^b, \\ \delta \mathbf{y} &= \mathbf{y}^o - [\mathcal{H}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x}^a - \mathbf{x}^b)], \\ &= [(\mathbf{y}^o - \mathbf{y}^t) + \mathbf{e}^H] - \mathbf{H}(\mathbf{x}^a - \mathbf{x}^t) = \mathbf{e}^o - \mathbf{H}\mathbf{e}^a. \end{aligned} \tag{A4}$$

Note that the observation error \mathbf{e}^o combines instrument and representativeness errors from the term $\mathbf{y}^o - \mathbf{y}^t$ and simulation errors from the term \mathbf{e}^H . Substitute Eq. (A4) into Eq. (A1) and rearrange to obtain

$$\mathbf{A}^{-1} \mathbf{e}^a = \mathbf{B}^{-1} \mathbf{e}^b + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{e}^o, \tag{A5}$$

where

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \tag{A6}$$

is a symmetric matrix. As will be shown, \mathbf{A} is the estimate of the analysis error covariance when the specified background and observation covariances are correct. Equation (A5) may be written as

$$\mathbf{e}^a = \mathbf{G}\mathbf{e}^b + \mathbf{K}\mathbf{e}^o, \tag{A7}$$

where

$$\begin{aligned} \mathbf{K} &= \mathbf{A}\mathbf{H}^T \mathbf{R}^{-1} \\ &= (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \\ &= \mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}, \end{aligned} \tag{A8}$$

and

$$\mathbf{G} = \mathbf{A}\mathbf{B}^{-1} = \mathbf{I} - \mathbf{K}\mathbf{H}. \tag{A9}$$

The second form of \mathbf{K} is the variational formulation, and the third is the optimal interpolation (OI) formulation. The equivalence of these two forms follows as a variant of the Woodbury matrix identity [see Kalnay (2002), Eq. (5.5.11)]. The Woodbury identity may also be used to show that $\mathbf{A}\mathbf{B}^{-1} = \mathbf{G}$.^{A1} According to Eq. (A7), the analysis error is partly due to the background error and partly due to the observation error. The first term is the background error updated by the Kalman filter covariance update matrix ($\mathbf{G} = \mathbf{I} - \mathbf{K}\mathbf{H}$). The second term is the observation error filtered by the Kalman filter gain matrix \mathbf{K} .

To obtain an expression for the true analysis error covariance $\hat{\mathbf{A}} = \langle \mathbf{e}^a \mathbf{e}^{aT} \rangle$, multiply Eq. (A7) by its transpose, take the expectation, and assume that \mathbf{e}^o and \mathbf{e}^b are independent in the sense that $\langle \mathbf{e}^o \mathbf{e}^{bT} \rangle$ vanishes. Then,

$$\langle \mathbf{e}^a \mathbf{e}^{aT} \rangle = \mathbf{G} \langle \mathbf{e}^b \mathbf{e}^{bT} \rangle \mathbf{G}^T + \mathbf{K} \langle \mathbf{e}^o \mathbf{e}^{oT} \rangle \mathbf{K}^T \tag{A10}$$

or

$$\hat{\mathbf{A}} = \mathbf{G}\hat{\mathbf{B}}\mathbf{G}^T + \mathbf{K}\hat{\mathbf{R}}\mathbf{K}^T = \hat{\mathbf{A}}^b + \hat{\mathbf{A}}^o, \tag{A11}$$

which is Eq. (3).

In the special case that $\mathbf{R} = \hat{\mathbf{R}}$ and $\mathbf{B} = \hat{\mathbf{B}}$, the entire first r.h.s. of Eq. (A11) reduces to \mathbf{A} and therefore $\hat{\mathbf{A}} = \mathbf{A}$.^{A2} In other words \mathbf{A} is the equal to the true analysis error covariance when the estimated background and observation error covariances are correctly specified.

Once $\hat{\mathbf{A}}$ is determined from Eq. (A11), its diagonal contains the estimated squared analysis error at the analysis locations:

$$\Delta_A = \text{diag}(\hat{\mathbf{A}}) = \text{diag}(\hat{\mathbf{A}}^b) + \text{diag}(\hat{\mathbf{A}}^o) = \Delta_A^b + \Delta_A^o. \tag{A12}$$

^{A1}Applying the Woodbury identity to \mathbf{A} gives $\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = \mathbf{B} - \mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}$. Thus, $\mathbf{A}\mathbf{B}^{-1} = \mathbf{I} - \mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \mathbf{H} = \mathbf{I} - \mathbf{K}\mathbf{H} = \mathbf{G}$.

^{A2}Replace \mathbf{G} and \mathbf{K} in Eq. (A11) with their original definitions $\mathbf{A}\mathbf{B}^{-1}$ and $\mathbf{A}\mathbf{H}^T \mathbf{R}^{-1}$ to obtain $\hat{\mathbf{A}} = \mathbf{A}(\mathbf{B}^{-1} \hat{\mathbf{B}}\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \hat{\mathbf{R}} \mathbf{R}^{-1} \mathbf{H})\mathbf{A}$, which, in the special case that $\mathbf{R} = \hat{\mathbf{R}}$ and $\mathbf{B} = \hat{\mathbf{B}}$, becomes $\hat{\mathbf{A}} = \mathbf{A}(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})\mathbf{A} = \mathbf{A}\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}$.

Here, Δ_A^b and Δ_A^o will be termed the background and observation squared analysis error, respectively. The mean square estimated analysis error and its components are similarly defined by Eq. (6) in terms of the trace of $\hat{\mathbf{A}}$.

REFERENCES

- Abdelnur Ruggiero, G., E. Cosme, J.-M. Brankart, J. L. Sommer, and C. Ubelmann, 2016: An efficient way to account for observation error correlations in the assimilation of data from the future SWOT high-resolution altimeter mission. *J. Atmos. Oceanic Technol.*, **33**, 2755–2768, <https://doi.org/10.1175/JTECH-D-16-0048.1>.
- Bauer, P., R. Buizza, C. Cardinali, and J. N. Thépaut, 2011: Impact of singular-vector-based satellite data thinning on NWP. *Quart. J. Roy. Meteor. Soc.*, **137**, 286–302, <https://doi.org/10.1002/qj.733>.
- Beal, D., P. Brasseur, J. Brankart, Y. Ourmières, and J. Verron, 2010: Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: Implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.*, **6**, 247–262, <https://doi.org/10.5194/os-6-247-2010>.
- Benjamin, S. G., 1989: An isentropic meso- α -scale analysis system and its sensitivity to aircraft and surface observations. *Mon. Wea. Rev.*, **117**, 1586–1603, [https://doi.org/10.1175/1520-0493\(1989\)117<1586:AIMSAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1586:AIMSAS>2.0.CO;2).
- Bergman, K. H., and W. D. Bonner, 1976: Analysis error as a function of observation density for satellite temperature soundings with spatially correlated errors. *Mon. Wea. Rev.*, **104**, 1308–1316, [https://doi.org/10.1175/1520-0493\(1976\)104<1308:AEAAFO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1308:AEAAFO>2.0.CO;2).
- Campbell, W. F., E. A. Satterfield, B. Ruston, and N. L. Baker, 2017: Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Mon. Wea. Rev.*, **145**, 1019–1032, <https://doi.org/10.1175/MWR-D-16-0240.1>.
- Dando, M. L., A. J. Thorpe, and J. R. Eyre, 2007: The optimal density of atmospheric sounder observations in the Met Office NWP system. *Quart. J. Roy. Meteor. Soc.*, **133**, 1933–1943, <https://doi.org/10.1002/qj.175>.
- Dee, D. P., 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3323–3343, <https://doi.org/10.1256/qj.05.137>.
- , and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295, <https://doi.org/10.1002/qj.49712454512>.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- Hotta, D., E. Kalnay, Y. Ota, and T. Miyoshi, 2017: EFSR: Ensemble forecast sensitivity to observation error covariance. *Mon. Wea. Rev.*, **145**, 5015–5031, <https://doi.org/10.1175/MWR-D-17-0122.1>.
- Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, 364 pp.
- Lazarus, S. M., M. E. Splitt, M. D. Lueken, R. Ramachandran, X. Li, S. Movva, S. J. Graves, and B. T. Zavodsky, 2010: Evaluation of data reduction algorithms for real-time analysis. *Wea. Forecasting*, **25**, 837–851, <https://doi.org/10.1175/2010WAF2222296.1>.
- Li, H., E. Kalnay, T. Miyoshi, and C. M. Danforth, 2009: Accounting for model errors in ensemble data assimilation. *Mon. Wea. Rev.*, **137**, 3407–3419, <https://doi.org/10.1175/2009MWR2766.1>.
- Li, X., J. Zhu, Y. Xiao, and R. Wang, 2010: A model-based observation-thinning scheme for the assimilation of high-resolution SST in the shelf and coastal seas around China. *J. Atmos. Oceanic Technol.*, **27**, 1044–1058, <https://doi.org/10.1175/2010JTECHO709.1>.
- Liu, Z.-Q., and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, 1367–1386, <https://doi.org/10.1256/003590002320373337>.
- , and —, 2003: The potential of high-density observations for numerical weather prediction: A study with simulated observations. *Quart. J. Roy. Meteor. Soc.*, **129**, 3013–3035, <https://doi.org/10.1256/qj.02.170>.
- Miyoshi, T., and M. Kunii, 2012: Using AIRS retrievals in the WRF-LETKF system to improve regional numerical weather prediction. *Tellus*, **64A**, 18408, <https://doi.org/10.3402/tellusa.v64i0.18408>.
- Nakabayashi, A., and G. Ueno, 2017: An extension of the ensemble Kalman filter for estimating the observation error covariance matrix based on the variational Bayes method. *Mon. Wea. Rev.*, **145**, 199–213, <https://doi.org/10.1175/MWR-D-16-0139.1>.
- Ochotta, T., C. Gebhardt, D. Saupe, and W. Wergen, 2005: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Quart. J. Roy. Meteor. Soc.*, **131**, 3427–3437, <https://doi.org/10.1256/qj.05.94>.
- Purser, R. J., D. F. Parrish, and M. Masutani, 2000: Meteorological observational data compression: An alternative to conventional “super-obbing.” NCEP Office Note 430, 12 pp., <http://www.emc.ncep.noaa.gov/mmb/papers/purser/on430.pdf>.
- Ramachandran, R., X. Li, S. Movva, S. Graves, S. Greco, D. Emmitt, J. Terry, and R. Atlas, 2005: Intelligent data thinning algorithm for earth system numerical model research and application. *21st Int. Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, San Diego, CA, Amer. Meteor. Soc., 13.8, https://ams.confex.com/ams/Annual2005/techprogram/paper_85987.htm.
- Richman, M. B., L. M. Leslie, T. B. Trafalis, and H. Mansour, 2015: Data selection using support vector regression. *Adv. Atmos. Sci.*, **32**, 277–286, <https://doi.org/10.1007/s00376-014-4072-9>.
- Todling, R., 2015a: A lag-1 smoother approach to system-error estimation: Sequential method. *Quart. J. Roy. Meteor. Soc.*, **141**, 1502–1513, <https://doi.org/10.1002/qj.2460>.
- , 2015b: A complementary note to ‘A lag-1 smoother approach to system-error estimation’: The intrinsic limitations of residual diagnostics. *Quart. J. Roy. Meteor. Soc.*, **141**, 2917–2922, <https://doi.org/10.1002/qj.2546>.
- Waller, J. A., D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard, 2016: Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Wea. Rev.*, **144**, 3533–3551, <https://doi.org/10.1175/MWR-D-15-0340.1>.
- Xu, Q., 2011: Measuring information content from observations for data assimilation: Spectral formulations and their implications

- to observational data compression. *Tellus*, **63A**, 793–804, <https://doi.org/10.1111/j.1600-0870.2011.00524.x>.
- , and L. Wei, 2011: Measuring information content from observations for data assimilations: Utilities of spectral formulations demonstrated with radar observations. *Tellus*, **63A**, 1014–1027, <https://doi.org/10.1111/j.1600-0870.2011.00542.x>.
- Zhu, T., and S. Boukabara, 2015: Development and test of the community satellite data thinning and representation optimization tool. *13th JCSDA Technical Review Meeting and Science Workshop on Satellite Data Assimilation*, College Park, MD, Joint Center for Satellite Data Assimilation, https://www.jcsda.noaa.gov/meetings_Wkshp2015_posters.php.